

Introduction to Learning and Intelligent Systems - Spring 2015

Martin Ivanov (ivanovma@student.ethz.ch)
Can Tuerk (can.tuerk@juniors.ethz.ch)
Jens Hauser(jhauser@student.ethz.ch)

April 26, 2015

Project 3 : Image Classification

Problem description

We received datafiles of images which we had to classify into ten different categories. The data consists of 2048 features and the amount of training data points is 40.000.

Solution

Our solution with a best performance of about 0.19 on the validation data due to the given loss function is mainly based on the following approach:

- To reduce the dimensions of our training data we first used a Principal Component approach with a randomized Singular Value Decomposition which keeps only the most significant singular vectors for the projection.
- After that we piped the reduced data to an gaussian Support Vector Classifier.

The best performance result we got with this approach is based on 500 *PrincipalComponents* and a penalty parameter C of 10 of the SVM.

Different Approaches

Unfortunately we didn't reach the hard baseline this time, although we spent a huge amount of time and tried the following approaches with a lot of different hyper parameters:

- ExtraTreesClassifier
- PCA - ExtraTreesClassifier
- RandomizedPCA - ExtraTreesClassifier
- kernelPCA - ExtraTreesClassifier
- PCA - LinearSVC
- kernelPCA - LinearSVC
- SGDClassifier
- RandomizedPCA - LinearSVC
- PCA - SGDClassifier
- RandomizedPCA - SGDClassifier
- PCA - SVC
- RandomizedPCA - SVC
- kernelPCA - SVC
- Neural Network - LinearSVC
- Neural Network - SVC